

BCB744 Biostatistics – Theory Test (Version 4)

Total: 135 marks | Time: 180 minutes

A. J. Smit
University of the Western Cape

2026-04-22

! Important

Instructions

- This paper has **three parts**: Part A (General Theory, 61 marks), Part B (Experiment Design and Hypothesis Formulation, 37 marks), and Part C (Statistical Output Interpretation, 37 marks).
- Mark allocations are shown next to each question in (/ **marks**) notation.
- Answer **all** questions.
- Write clearly and in complete sentences where prose is required.
- Number all questions clearly and use the Quarto headings facility to assign the main question number to level 1 (e.g., # **Question 1**) and the subordinate parts to level 2 (e.g., ## **Q1a**).
- Statistical notation: use H_0 for the null hypothesis and H_A for the alternative hypothesis.
- You are **not** allowed access to the internet or AI.
- You **may** use the cheatsheet and the RStudio/R help files.
- You **must** submit your knitted document in `.html` format on iKamva immediately after the 3-hr test duration has elapsed.
- Use `embed-resources: true` in Quarto's YAML header to ensure the `.html` file displays correctly.
- **Any** format other than `.html` will be disqualified from assessment.

Part A: General Theory (61 marks)

Question 1 – Observational vs Experimental Studies (/5)

- a. What is the fundamental distinction between an **observational** study and a **controlled experiment**? (/ 2)
- b. Why is it generally not possible to draw causal conclusions from an observational study, even when a strong statistical association exists? (/ 2)

- c. Give **one** example from biological sciences where an observational study is the only ethical or practical option. (/ 1)

 Tip

Model Answer – Question 1

a.

- ✓ In a **controlled experiment**, the researcher deliberately manipulates one or more factors (independent variables), randomly assigns subjects to treatment conditions, and controls all other variables – thereby allowing cause-and-effect inferences.
- ✓ In an **observational study**, the researcher measures variables as they naturally occur, without manipulation or random assignment – associations can be detected but the researcher cannot isolate the causal factor.

b.

- ✓ Without random assignment to treatment groups, subjects that differ in the predictor variable (e.g., smokers vs. non-smokers) may also differ in many other ways (confounding variables). Any observed association could be driven by these unmeasured differences rather than the predictor itself.
- ✓ Additionally, the direction of causality is ambiguous: the putative “cause” and “effect” may share a common upstream driver, or the arrow of causation may run in the opposite direction from what is assumed.

c.

- ✓ Any one of: studying the health effects of smoking in humans (unethical to randomly assign people to smoke); tracking the long-term effects of a natural disaster on wildlife populations; observing migration routes or breeding behaviour of endangered species that cannot be disturbed.

Question 2 – Variables and Study Design (/7)

- a. Distinguish between a **continuous** and a **categorical** (nominal) predictor variable. Give one biological example of each. (/ 3)
- b. What is the difference between a **fixed factor** and a **random factor** in an experimental design? Give one example of each from a field ecology context. (/ 2)
- c. Define **pseudoreplication**. Why is it statistically problematic? (/ 2)

💡 Tip

Model Answer – Question 2

a.

- ✓ A **continuous predictor** takes any real value on a scale (e.g., water temperature in °C, salinity in ppt, or body mass in grams).
- ✓ A **categorical predictor** groups observations into discrete, named categories with no inherent numeric spacing (e.g., species identity, treatment group, sex, or habitat type).
- ✓ Biological examples: continuous – seawater pH as a predictor of coral calcification rate; categorical – diet type (herbivore, omnivore, carnivore) as a predictor of gut passage time.

b.

- ✓ A **fixed factor** is one whose specific levels are deliberately chosen and are the levels of direct interest (conclusions apply only to those levels). Example: three specific herbicide concentrations selected by the researcher.
- ✓ A **random factor** is one whose levels are a random sample from a broader population of possible levels, and the goal is to generalise to that population. Example: 10 randomly chosen study plots within a forest biome, representing all possible plots.

c.

- ✓ Pseudoreplication occurs when multiple measurements from the same experimental unit (or non-independent observations) are treated as independent replicates, artificially inflating the apparent sample size.
- ✓ This is statistically problematic because it underestimates the true variance within treatments, inflates the *F*- or *t*-statistic, and dramatically inflates the Type I error rate – leading to false conclusions of significance.

Question 3 – Data Visualisation (/5)

- a. When is a **boxplot** more informative than a bar chart showing mean \pm standard error? (/ 2)
- b. What feature of a dataset does a **violin plot** reveal that a boxplot does not? (/ 1)
- c. Identify the five summary statistics encoded in a standard boxplot (Tukey box-and-whisker). (/ 2)

💡 Tip

Model Answer – Question 3

a.

- ✓ A boxplot is more informative when the data are skewed or contain outliers – it directly shows the median, spread, skewness (via asymmetric box), and outlier positions that a mean \pm SE bar chart obscures.
- ✓ It is particularly useful when sample sizes are modest and the mean is not a robust summary (e.g., bimodal or heavy-tailed distributions), or when comparing distributions that differ in shape rather than just location.

b.

- ✓ A violin plot reveals the **full probability density** (distributional shape) of the data – including multimodality, skew, and the concentration of values at different levels – something a boxplot cannot show since it only summarises five quantiles.

c.

- ✓✓ The five summary statistics: **lower quartile** (Q1, 25th percentile), **median** (Q2, 50th percentile), **upper quartile** (Q3, 75th percentile), and the two **whisker endpoints** that extend to the most extreme values within $1.5 \times$ IQR of Q1 and Q3. Points beyond the whiskers are plotted individually as **outliers**.

Question 4 – ANOVA and Post-hoc Tests (/6)

- a. Explain why it is statistically incorrect to perform all pairwise comparisons between three or more groups using individual t -tests, rather than ANOVA. (/ 3)
- b. What is the **Tukey Honestly Significant Difference (HSD)** test? When is it the appropriate post-hoc procedure following a significant one-way ANOVA result? (/ 3)

💡 Tip

Model Answer – Question 4

a.

- ✓ Each individual t -test is conducted at $\alpha = 0.05$, meaning there is a 5% chance of a Type I error per test. With k groups, there are $k(k-1)/2$ pairwise comparisons: for $k = 3$, that is 3 comparisons; for $k = 5$, that is 10.
- ✓ The **family-wise error rate** (FWER) – the probability of making *at least one* false rejection across all tests – inflates substantially: with 3 independent tests, $\text{FWER} \approx 1 - (0.95)^3 \approx 0.14$, not 0.05. With 10 tests, $\text{FWER} \approx 0.40$.
- ✓ ANOVA conducts a single omnibus F -test that controls the error rate at $\alpha = 0.05$ for the global null hypothesis (all means equal), avoiding this inflation.

b.

- ✓ The Tukey HSD test is a **post-hoc multiple comparison procedure** that makes all pairwise comparisons among group means while controlling the family-wise error rate at α across all comparisons. It uses the studentised range distribution to compute critical differences.
- ✓ It is appropriate when: (a) the omnibus ANOVA F -test is significant (indicating *some* difference exists), (b) all groups have approximately equal sample sizes (balanced or near-balanced design), and (c) the researcher wants to identify *which specific pairs* of groups differ significantly, with simultaneous Type I error control across all pairwise tests.

Question 5 – Correlation vs Causation (/7)

- a. Explain the conceptual difference between **correlation analysis** and **simple linear regression**, even though both describe relationships between two continuous variables. (/ 3)
- b. A researcher reports a correlation of $r = 0.85$ ($p < 0.001$) between ocean surface temperature ($^{\circ}\text{C}$) and the frequency of harmful algal bloom (HAB) events per year, based on a 25-year observational time series. A journalist headlines this as “Warming Oceans Proven to Trigger Algal Blooms.” Identify **two** alternative explanations for this correlation and explain why the journalist’s conclusion is premature. (/ 4)



Tip

Model Answer – Question 5

a.

- ✓ **Correlation** quantifies the *strength and direction* of the linear association between two variables, treating both symmetrically – there is no distinction between predictor and response. Pearson's r ranges from -1 to $+1$ and is scale-free.
- ✓ **Simple linear regression** explicitly models one variable (the *response*, y) as a function of the other (the *predictor*, x), estimating the slope (rate of change of y per unit x) and intercept. It is used for prediction and for quantifying the *magnitude* of the relationship.
- ✓ Regression imposes an asymmetric causal framework (x influences y) and provides a fitted equation; correlation does not imply or require directionality and only describes the co-variation.

b.

- ✓ **Shared temporal trend (spurious correlation)**: both SST and HAB events may be independently increasing over time due to long-term climate change and increases in coastal nutrient loading (eutrophication), respectively. The time series correlation captures their shared temporal trajectory, not a direct mechanistic link.
- ✓ **Confounding by coastal nutrient enrichment**: warmer years may coincide with drier years that reduce river flushing of coastal waters, concentrating nutrients and promoting blooms – it is nutrient concentration that directly causes blooms, not temperature per se.
- ✓ **Why the conclusion is premature**: correlation establishes association, not causation. The direction of influence cannot be confirmed from observational data alone; no manipulation of SST was performed, and multiple confounders and alternative mechanisms have not been ruled out. Establishing causation requires controlled experiments, mechanistic pathway confirmation, or at minimum a structural causal model with confounders accounted for.

(Accept any two distinct alternatives, 2 marks each; 2 marks for the methodological explanation.)

Question 6 – t -tests (/7)

- a. Explain the conceptual difference between an **independent-samples** t -test and a **paired** t -test. Give a biological scenario where the paired design is more appropriate. (/ 3)
- b. Under what conditions should you use **Welch's t -test** rather than Student's t -test? (/ 2)
- c. What non-parametric test is the appropriate alternative when the assumptions of the independent-samples t -test cannot be met? (/ 2)



Tip

Model Answer – Question 6

a.

- ✓ An **independent-samples t -test** compares means from two separate, unrelated groups – the observations in one group have no pairing or correspondence with observations in the other.
- ✓ A **paired t -test** compares means where each observation in one condition is logically matched to an observation in the other – differences are computed within pairs before testing, removing between-individual variation.
- ✓ Biological scenario: measuring cortisol in the same individual fish *before* and *after* a stressor. Because both measurements come from the same fish, pairing removes individual-level baseline differences and increases power to detect the stress response.

b.

- ✓ Welch's t -test should be used when the two groups have **unequal variances** (heteroscedasticity), as it adjusts the degrees of freedom to account for this.
- ✓ It is generally recommended as the default over Student's t -test even when variances appear similar, because it is valid under both equal and unequal variances with minimal power cost when variances happen to be equal.

c.

- ✓✓ The **Wilcoxon rank-sum test** (also called the Mann-Whitney U test) is the appropriate non-parametric alternative when normality or other assumptions of the independent-samples t -test are violated. It compares the distributions of two independent groups using ranked data.

Question 7 – Residual Diagnostics (/7)

Residual diagnostic plots are central tools for evaluating whether a fitted model meets its underlying assumptions. Describe what each of the following patterns in diagnostic plots suggests, what assumption is violated, and what corrective action you might take.

- a. A **fan-shaped** (heteroscedastic) pattern in the residuals-vs-fitted plot, where residuals widen as fitted values increase. (/ 2)
- b. A systematic **S-shaped curve** in the normal Q-Q plot of residuals. (/ 2)
- c. A **U-shaped (concave) curve** in the residuals-vs-fitted plot. (/ 2)
- d. One or two points with very large residuals far from the main point cloud. (/ 1)

 Tip

Model Answer – Question 7

a.

- ✓ A fan shape indicates **heteroscedasticity** – the residual variance is not constant but increases with the fitted values. This violates the homoscedasticity assumption.
- ✓ Corrective actions: apply a variance-stabilising transformation to the response (e.g., log or square-root); fit a weighted least squares model; or use a generalised linear model with an appropriate error family (e.g., Poisson or Gamma with a log link).

b.

- ✓ An S-shaped curve in the Q-Q plot indicates **heavy tails** (leptokurtosis) if the S bends upward on the right and downward on the left – residuals are more extreme than expected under normality. The **normality** assumption is violated.
- ✓ Corrective actions: consider a transformation (e.g., log or Box-Cox); investigate whether the extreme residuals correspond to outliers that should be examined; use a robust regression method or a distribution with heavier tails (e.g., t-distribution errors).

c.

- ✓ A U-shaped (or arch-shaped) curve in the residuals-vs-fitted plot indicates **non-linearity** – the fitted linear model systematically under- or over-predicts in different regions of the predictor space.
- ✓ Corrective actions: add a polynomial (quadratic) term to the model; apply a transformation to the predictor variable; consider a non-linear or generalised additive model (GAM).

d.

- ✓ Large isolated residuals indicate potential **outliers** or **influential observations** – individual data points that deviate markedly from the model's predictions. They may represent data entry errors, genuinely unusual biological events, or evidence that the model is misspecified for a subset of the data. These should be investigated (not automatically removed) by examining the raw data and leverage/influence statistics (Cook's distance).

Question 8 – Selecting the Correct Statistical Test (/7)

- a. You are comparing body condition index (a continuous measure) across **three groups** of an endangered antelope species (wet-season migrants, dry-season migrants, and year-round residents). A Shapiro-Wilk test on each group shows the data are not normally distributed. What test would you use, and provide **three** reasons for your choice. (/ 4)
- b. You have data on the feeding rates of 15 individual fish, measured once under ambient light and once under reduced light. What information would you need to confirm before deciding between an independent-samples *t*-test and a paired *t*-test? (/ 3)

💡 Tip

Model Answer – Question 8

a.

- ✓ **Kruskal-Wallis rank-sum test** (the non-parametric analogue of one-way ANOVA).
- ✓ Reason 1: There are **three independent groups**, requiring a test that can simultaneously compare more than two groups (ruling out pairwise *t*-tests or Wilcoxon rank-sum, which are two-sample methods).
- ✓ Reason 2: The data are **not normally distributed** within groups (Shapiro-Wilk significant), violating a key assumption of one-way ANOVA. Kruskal-Wallis operates on ranks and requires no distributional assumption.
- ✓ Reason 3: The response (body condition index) is continuous and the groups are independent (different individual animals), meeting the design requirements for Kruskal-Wallis.

b.

- ✓ You need to confirm whether the same 15 individual fish were measured under **both** light conditions (repeated measures / within-subject design) or whether the 30 measurements come from **different** fish assigned to each condition (between-subject design).
- ✓ If the same fish were measured twice (paired), a paired *t*-test (or Wilcoxon signed-rank if non-normal) is appropriate and more powerful.
- ✓ If different fish were used for each condition, an independent-samples *t*-test (or Wilcoxon rank-sum) is correct. Applying a paired test to unpaired data, or vice versa, is a design error that can produce invalid results.

Question 9 – Multiple Regression and Interaction Effects (/10)

- a. What is **multicollinearity** in a multiple regression context, and how does it affect the interpretation of regression coefficients? (/ 3)
- b. What is the **Variance Inflation Factor (VIF)**, and what threshold is commonly used to flag problematic multicollinearity? (/ 2)
- c. A researcher fits the following model to data on algal growth rate (cm day⁻¹):
`growth ~ temperature + nutrient + temperature:nutrient`. The interaction term `temperature:nutrient` is significant. Explain what a significant interaction means for the *interpretation* of the main effect of temperature. How would you interpret the interaction coefficient in practical, biological terms? (/ 5)



Tip

Model Answer – Question 9

a.

- ✓ **Multicollinearity** occurs when two or more predictors in a multiple regression model are highly correlated with each other – they share much of the same variance in the response.
- ✓ Consequence: the individual regression coefficients become **unstable** (high standard errors), because the model cannot reliably partition the variation in the response between the collinear predictors. Small changes in the dataset can produce large swings in coefficient estimates.
- ✓ The *combined* predictive power of the model may remain unaffected, but the *individual* coefficients can no longer be interpreted as the effect of one variable holding the other constant – because in reality the predictors cannot be independently varied.

b.

- ✓ The **VIF** for predictor j is $1 / (1 - R_j^2)$, where R_j^2 is the proportion of variance in predictor j explained by all other predictors. It quantifies how much the variance of the coefficient estimate is inflated by collinearity.
- ✓ A common threshold: $VIF > 5$ (some authorities use 10) is flagged as problematic; $VIF = 1$ indicates no collinearity.

c.

- ✓ When the interaction term is significant, the main effect of temperature is **conditional** – it is not a single universal effect but depends on the level of the nutrient variable. The “main effect” label in the coefficient table describes the effect of temperature only when nutrients are at the reference level (e.g., ambient), not across all nutrient conditions.
- ✓ The interaction coefficient represents the *additional change in the slope of temperature* when nutrients change from the reference level (e.g., ambient) to the comparison level (e.g., enriched). In other words, the temperature effect on growth rate differs between ambient and enriched nutrient conditions.
- ✓ Biological example: if the interaction coefficient is positive (+0.12), the growth rate increases by an additional 0.12 cm day⁻¹ per °C of warming under enriched nutrients compared to ambient nutrients. This means warming has a **stronger stimulatory effect on growth when nutrients are not limiting** – temperature and nutrient availability operate synergistically, not independently.
- ✓ To fully describe the relationship, you must report **separate slope estimates** for each nutrient level (the conditional effects), rather than a single temperature effect, because the single “main effect” is misleading when the interaction is significant.
- ✓ This has important implications for biological interpretation: if nutrient enrichment amplifies the warming response, then eutrophication and ocean warming may act synergistically to increase macroalgal proliferation – a non-additive interaction that cannot be predicted from single-factor studies.

Part B: Experiment Design and Hypothesis Formulation (37 marks)

Question 10 – Factorial Design: Lizard Sprint Speed (/13)

A herpetologist measures the maximum sprint speed (m s^{-1}) of common lizards (*Zootoca vivipara*) reared under two temperatures (20°C and 30°C) and two diet types (insect-based and plant-based). Six individuals are assigned to each of the four treatment combinations. The first six rows of the dataset are:

| | lizard_id | temperature | diet_type | sprint_speed_m_s |
|---|-----------|-------------|------------|------------------|
| 1 | 1 | 20°C | insects | 1.23 |
| 2 | 2 | 20°C | insects | 1.18 |
| 3 | 3 | 20°C | vegetation | 0.89 |
| 4 | 4 | 20°C | vegetation | 0.92 |
| 5 | 5 | 30°C | insects | 1.67 |
| 6 | 6 | 30°C | insects | 1.71 |

The researcher asks: “Does sprint speed vary with temperature, diet type, or the interaction between them?”

- State formal null and alternative hypotheses for each of the following effects: (i) the main effect of temperature, (ii) the main effect of diet type, and (iii) the temperature \times diet interaction. (/ 6)
- What statistical test is most appropriate, and give **three** reasons, including reference to the number of predictors and their nature. (/ 4)
- The temperature \times diet interaction is significant. What does this mean biologically? How does it affect how you would report and interpret the main effects? (/ 3)

💡 Tip

Model Answer – Question 10

a. Two marks per effect pair ($H_0 + H_A$):

(i) *Temperature:*

- ✓ H_0 : Mean sprint speed does not differ between lizards reared at 20°C and 30°C ($\mu_{20} = \mu_{30}$).
- ✓ H_A : Mean sprint speed differs between the two temperature treatments ($\mu_{20} \neq \mu_{30}$).

(ii) *Diet type:*

- ✓ H_0 : Mean sprint speed does not differ between lizards fed insects and those fed vegetation ($\mu_{\text{insects}} = \mu_{\text{vegetation}}$).
- ✓ H_A : Mean sprint speed differs between the two diet types ($\mu_{\text{insects}} \neq \mu_{\text{vegetation}}$).

(iii) *Temperature × diet interaction:*

- ✓ H_0 : The effect of temperature on sprint speed is the same regardless of diet type (no interaction; the effects are additive).
- ✓ H_A : The effect of temperature on sprint speed depends on diet type (the two factors interact; their combined effect is not simply additive).

b.

- ✓ **Two-way (factorial) ANOVA** – this is the correct test because there are two categorical predictors (temperature with 2 levels; diet with 2 levels) and a single continuous response variable (sprint speed).
- ✓ Reason 1: There are **two factorial predictors** (not one), each with distinct levels. A two-way ANOVA simultaneously tests main effects of each factor and their interaction – a design that one-way ANOVA or *t*-tests cannot accommodate.
- ✓ Reason 2: The response (sprint speed, m s^{-1}) is **continuous** and ratio-scaled, appropriate for ANOVA which compares group means.
- ✓ Reason 3: The design is **balanced** (equal replication, 6 per cell), which maximises the power and interpretive clarity of a factorial ANOVA; each cell's mean is estimated with equal precision.

c.

- ✓ A significant interaction means that the effect of temperature on sprint speed **depends on diet type** (or equivalently, the diet effect depends on temperature). The two factors do not act independently.
- ✓ For example, warming may strongly enhance sprint speed in insect-fed lizards (because sufficient protein supports muscle development) but have little effect in vegetation-fed lizards (because plant-based nutrition cannot support the thermal enhancement of locomotor performance).
- ✓ Because the interaction is significant, the **main effects cannot be interpreted in isolation** – reporting a single main effect of temperature (e.g., “warmer lizards are faster”) is misleading if this is only true for one diet type. You must present and interpret the **conditional effects** (simple main effects) separately for each diet type, ideally via an interaction plot.

Question 11 – Bacterial Colony Counts Across Antibiotic Concentrations (/12)

A microbiologist grows *Staphylococcus aureus* at four concentrations of a novel antibiotic (0, 10, 50, and 100 $\mu\text{g mL}^{-1}$) with three replicate cultures per concentration. Colony counts (CFU mL^{-1}) are recorded after 24 hours. The first eight rows of the dataset are:

| | replicate | conc_ug_mL | colony_CFU_mL |
|---|-----------|------------|---------------|
| 1 | 1 | 0 | 4500 |
| 2 | 2 | 0 | 5120 |
| 3 | 3 | 0 | 4800 |
| 4 | 1 | 10 | 1230 |
| 5 | 2 | 10 | 980 |
| 6 | 3 | 10 | 1105 |
| 7 | 1 | 50 | 213 |
| 8 | 2 | 50 | 178 |

The research question is: “Does antibiotic concentration significantly affect bacterial colony count?”

- Formulate formal null and alternative hypotheses. (/ 3)
- Identify two appropriate statistical tests you might apply to these data, explaining with **specific reference** to the nature of the response variable and the experimental design. (/ 6)
- What transformation might make the data more amenable to a parametric test, and what property would it stabilise? (/ 3)

Question 12 – Cortisol Response to Acute Stress in Zebrafish (/12)

A researcher measures plasma cortisol (ng mL^{-1}) in 20 individual zebrafish (*Danio rerio*) before and after a 5-minute confinement stressor. The same 20 fish are measured at both time points. The first six rows of the dataset are:

| | fish_id | timepoint | cortisol_ng_mL |
|---|---------|-----------|----------------|
| 1 | 1 | before | 12.4 |
| 2 | 2 | before | 15.1 |
| 3 | 3 | before | 11.8 |
| 4 | 1 | after | 28.7 |
| 5 | 2 | after | 31.2 |
| 6 | 3 | after | 25.9 |

The research question is: “Does acute confinement stress significantly increase plasma cortisol in zebrafish?”

- Formulate appropriate null and alternative hypotheses. Note whether the alternative is directional (one-tailed) or non-directional (two-tailed) and justify this choice. (/ 3)
- Identify the most appropriate statistical test and give **three** reasons for your choice, with reference to the data structure. (/ 5)
- What aspect of normality would you check for this test specifically, and how would you check it? (/ 2)
- If the result is statistically significant, what additional information would you report to describe the biological magnitude of the effect? (/ 2)

💡 Tip

Model Answer – Question 12

a.

- ✓ H_0 : The mean cortisol level does not change following confinement stress; the mean difference (after – before) = 0.
- ✓ H_A : The mean cortisol level is significantly **higher** after confinement stress than before; mean difference > 0. This is a **one-tailed** alternative.
- ✓ A one-tailed test is justified here because the HPA/HPI axis stress physiology of teleost fish is well-established: confinement is known to activate cortisol release. A directional prediction is supported by strong *a priori* mechanistic knowledge, not by inspecting the data.

b.

- ✓ **Paired *t*-test** (or Wilcoxon signed-rank test if the paired differences are non-normal).
- ✓ Reason 1: The **same individuals** are measured at both time points – each before-measurement is intrinsically linked to the after-measurement from the same fish. This is a within-subject (repeated-measures) design, requiring a paired analysis.
- ✓ Reason 2: Using an independent-samples *t*-test would ignore this pairing, increasing residual variance (between-fish baseline differences) and reducing power.
- ✓ Reason 3: The response variable (cortisol, ng mL⁻¹) is continuous and ratio-scaled, meeting the measurement-scale requirement for a parametric test.

c.

- ✓ For a paired *t*-test, the assumption of normality applies to the **paired differences** (after – before), not to the raw cortisol values. Calculate each difference (e.g., 28.7 – 12.4 = 16.3), then apply a Shapiro-Wilk test or examine a Q-Q plot of the 20 difference values.
- ✓ If the differences are non-normal (especially with only 20 pairs), the **Wilcoxon signed-rank test** is the appropriate non-parametric alternative.

d.

- ✓ Report the **mean (or median) difference** in cortisol (e.g., mean increase of X ng mL⁻¹) alongside its 95% confidence interval, to convey the biological magnitude of the stress response.
- ✓ Cohen's *d* (effect size) could also be reported, comparing the mean difference to the SD of the differences, to indicate whether the magnitude is small, medium, or large in practical terms.

Part C: Statistical Output Interpretation (37 marks)

Question 13 – Simple Linear Regression Summary (/12)

A researcher models the relationship between soil salinity (ppt) and plant height (cm) in *S. alterniflora*. The `lm()` output is:

```
Call:
lm(formula = height_cm ~ salinity_ppt, data = spartina)

Residuals:
    Min       1Q   Median       3Q      Max
-6.312  -1.874   0.102   1.993   5.876

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.8312     1.4231   45.55 < 2e-16 ***
salinity_ppt  -0.8847     0.0512  -17.28 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.641 on 48 degrees of freedom
Multiple R-squared:  0.8614, Adjusted R-squared:  0.8585
F-statistic: 298.6 on 1 and 48 DF, p-value: < 2.2e-16
```

- Write the equation of the fitted regression line. (/ 2)
- Interpret the slope coefficient (-0.8847) in biological terms. (/ 2)
- What does $R^2 = 0.8614$ mean? Why is the adjusted R^2 (0.8585) slightly lower? (/ 3)
- The residuals range from -6.312 to 5.876 . What do these values represent, and are there any obvious concerns from this summary? (/ 3)
- What do the *** significance codes next to both coefficients indicate, and what is the p -value threshold they correspond to? (/ 2)

💡 Tip

Model Answer – Question 13

a.

- ✓ ✓ $\widehat{height} = 64.831 - 0.885 \times salinity_ppt$

(Accept coefficient values rounded to 2–3 decimal places.)

b.

- ✓ For every 1 ppt increase in soil salinity, plant height is predicted to **decrease** by approximately 0.88 cm.
- ✓ This negative slope is consistent with the hypothesis that saline stress limits shoot elongation – plants in saltier soils are reliably shorter than those in fresher conditions.

c.

- ✓ $R^2 = 0.8614$ means that **86.14% of the total variation in plant height** is explained by variation in soil salinity. The model accounts for the vast majority of height differences among plants.
- ✓ The adjusted R^2 (0.8585) is slightly lower because it penalises the model for the number of predictors relative to sample size – with only one predictor and 50 observations the penalty is small, but adjusted R^2 always decreases relative to R^2 to avoid inflated estimates from adding irrelevant predictors.

d.

- ✓ The residuals are the differences between each observed height and the height predicted by the model ($e_i = y_i - \hat{y}_i$). They represent variation in height that is *not* explained by salinity.
- ✓ The maximum residual (5.876 cm) and minimum (–6.312 cm) are moderately larger than a quarter of the data spread (IQR roughly –1.874 to 1.993), suggesting at least one or two individual plants deviate more strongly from the fitted line. This is not alarming but warrants checking the residuals-vs-fitted plot for any influential outliers or systematic patterns.
- ✓ There is no strong asymmetry between the Min (–6.312) and Max (5.876) residuals, suggesting broadly symmetric residuals around zero – consistent with the normality assumption.

e.

- ✓ *** indicates a p -value < 0.001 : the probability of observing a slope (or intercept) this large or larger under the null hypothesis is less than 0.1%.
- ✓ Both the intercept and the slope are highly significant – we have very strong evidence that the intercept differs from zero and that the slope of salinity on height is non-zero, i.e., that the relationship exists.

Question 14 – One-Way ANOVA with Tukey HSD Post-hoc (/12)

A researcher compares the maximum sprint speed (m s^{-1}) of lizards from four habitat types: Desert, Grassland, Forest, and Savanna. Ten lizards per habitat are measured. The ANOVA and Tukey HSD results are:

Analysis of Variance Table

Response: sprint_speed

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|------------|
| habitat | 3 | 4.821 | 1.607 | 12.44 | <0.001 *** |
| Residuals | 36 | 4.651 | 0.129 | | |

Tukey multiple comparisons of means
95% family-wise confidence level

\$habitat

| | diff | lwr | upr | p adj |
|-------------------|-------|--------|-------|--------|
| Forest-Desert | 0.312 | 0.089 | 0.535 | 0.0024 |
| Grassland-Desert | 0.187 | -0.036 | 0.410 | 0.1210 |
| Forest-Grassland | 0.125 | -0.098 | 0.348 | 0.4120 |
| Savanna-Desert | 0.528 | 0.305 | 0.751 | 0.0001 |
| Savanna-Forest | 0.216 | -0.007 | 0.439 | 0.0612 |
| Savanna-Grassland | 0.341 | 0.118 | 0.564 | 0.0008 |

- State the null hypothesis being tested by the ANOVA F -test. (/ 2)
- Interpret the F -value (12.44) and the associated p -value. What conclusion do you draw from the ANOVA alone? (/ 3)
- Based on the Tukey HSD output, identify all **significantly** and **non-significantly** different pairs of habitat types. (/ 4)
- The Tukey test uses a “95% family-wise confidence level.” What does this mean, and why is it preferable to performing all pairwise comparisons each at $\alpha = 0.05$? (/ 3)

💡 Tip

Model Answer – Question 14

a.

- ✓ H_0 : The mean sprint speed is equal across all four habitat types ($\mu_{\text{Desert}} = \mu_{\text{Grassland}} = \mu_{\text{Forest}} = \mu_{\text{Savanna}}$).
- ✓ H_A (implicit): At least one habitat type has a mean sprint speed that differs from the others.

b.

- ✓ $F(3, 36) = 12.44$ means that the between-group variance is 12.44 times larger than the within-group (residual) variance – the habitat groups differ far more than would be expected from random sampling of a common population.
- ✓ $p < 0.001 \ll \alpha = 0.05$: we reject H_0 . There is very strong statistical evidence that mean sprint speed differs among at least some habitat types.
- ✓ However, the ANOVA alone does not identify *which* habitats differ – only that differences exist. Post-hoc testing is required to pinpoint the specific pairwise differences.

c.

Significantly different pairs (adjusted $p < 0.05$):

- ✓ Forest vs. Desert ($p = 0.0024$) – Forest lizards sprint faster than Desert lizards.
- ✓ Savanna vs. Desert ($p = 0.0001$) – Savanna lizards sprint fastest vs. Desert.
- ✓ Savanna vs. Grassland ($p = 0.0008$) – Savanna differs from Grassland.

Non-significantly different pairs (adjusted $p > 0.05$):

- ✓ Grassland vs. Desert ($p = 0.1210$) – no significant difference.
- Forest vs. Grassland ($p = 0.4120$) – no significant difference.
- Savanna vs. Forest ($p = 0.0612$) – borderline, not significant at $\alpha = 0.05$.

(Award 1 mark per correctly classified pair, up to 4 marks total; accept minor omissions.)

d.

- ✓ “95% family-wise confidence level” means that there is a 95% probability that **all** confidence intervals in the table simultaneously contain the true pairwise differences – the error is controlled across the entire family of 6 comparisons, not separately per interval.
- ✓ If all 6 comparisons were each run at $\alpha = 0.05$, the family-wise Type I error rate would be approximately $1 - (0.95)^6 \approx 0.26$ – a 26% chance of at least one false positive among the six tests.
- ✓ Tukey HSD adjusts the critical difference threshold so that the *combined* probability of any false positive across all comparisons remains at 5%, providing rigorous control while remaining more powerful than simpler corrections (e.g., Bonferroni) when all pairwise comparisons are of interest.

Question 15 – Multiple Regression with an Interaction Term (/13)

An environmental physiologist models the growth rate (cm day^{-1}) of a marine macroalga as a function of seawater temperature (continuous, $^{\circ}\text{C}$) and nutrient level (categorical: Low vs. High). An interaction term is included. The `lm()` output is:

```
Call:
lm(formula = growth_rate ~ temperature + nutrient + temperature:nutrient,
    data = algae)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.2450     0.3870   -3.22  0.0020 **
temperature         0.1870     0.0310    6.03 < 0.001 ***
nutrientHigh       2.3410     0.4120    5.68 < 0.001 ***
temperature:nutrientHigh 0.1240     0.0520    2.38  0.0204 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.834 on 56 degrees of freedom
Multiple R-squared:  0.7891, Adjusted R-squared:  0.7769
F-statistic: 69.81 on 3 and 56 DF,  p-value: < 2.2e-16
```

- How many predictor variables are in this model (count distinct predictors, not rows in the table)? (/ 1)
- Interpret the coefficient for `temperature` (0.1870) in the context of this model. (/ 2)
- The interaction term (`temperature:nutrientHigh`) has a coefficient of 0.1240 and is significant. Explain what this means for the biological relationship being studied. (/ 4)
- Write out the full regression equation separately for (i) **Low-nutrient** algae and (ii) **High-nutrient** algae. (/ 3)
- What does the adjusted R^2 (0.7769) indicate, and why is it reported in preference to R^2 (0.7891)? (/ 3)

💡 Tip

Model Answer – Question 15

a.

- ✓ **Two** distinct predictor variables: temperature (continuous) and nutrient level (categorical, two levels). The interaction term is derived from these two and is not an additional independent predictor.

b.

- ✓ The coefficient 0.1870 for temperature represents the effect of temperature **for Low-nutrient algae** (the reference level of the nutrient factor): for each 1°C increase in temperature, growth rate increases by approximately 0.187 cm day⁻¹, when nutrients are at the Low level.
- ✓ This is the **conditional** slope – because an interaction term is present, this coefficient does not apply uniformly to all algae; it specifically describes the temperature effect under Low-nutrient conditions.

c.

- ✓ The significant interaction coefficient (0.1240) indicates that the **effect of temperature on growth rate is stronger under High-nutrient conditions** than under Low-nutrient conditions.
- ✓ Specifically: in High-nutrient conditions, the growth rate increases by an **additional** 0.124 cm day⁻¹ per °C compared to the Low-nutrient slope (0.1870), giving a combined temperature slope of 0.1870 + 0.1240 = 0.311 cm day⁻¹ per °C under High nutrients.
- ✓ Biologically: nutrients appear to be **co-limiting** with temperature. When nutrients are abundant, warming has a larger stimulatory effect on growth, possibly because the biochemical machinery for photosynthesis and protein synthesis can operate at higher rates when both thermal energy and building materials are available.
- ✓ This synergistic interaction means that in oligotrophic (nutrient-poor) systems, ocean warming will have a smaller effect on macroalgal growth than in eutrophic (nutrient-rich) coastal environments – an important distinction for management of coastal blooms under climate change.

d.

- ✓ (i) **Low-nutrient algae** (nutrientHigh = 0): $\widehat{growth} = -1.245 + 0.187 \times temperature$
- ✓✓ (ii) **High-nutrient algae** (nutrientHigh = 1; add both the nutrientHigh coefficient and the interaction term): $\widehat{growth} = (-1.245 + 2.341) + (0.187 + 0.124) \times temperature = 1.096 + 0.311 \times temperature$

e.

- ✓ Adjusted $R^2 = 0.7769$ means that approximately **77.7% of the variation in algal growth** is explained by the model (comparing $R^2(0.816)$ to the coefficient of determination for a model with no predictors). It is a good indicator of model fit that adjusts for model complexity; it increases when irrelevant predictors are added; it decreases if a predictor adds less explanatory power than expected by chance.

End of Version 4

Bibliography